

**161—8.1(216) General provisions—employee selection procedures.**

**8.1(1)** *“Test” defined.* For the purpose of the rules in this chapter, the term “test” is defined as any paper-and-pencil or performance measure used as a basis for any employment decision. The rules in this chapter apply, for example, to ability tests which are designed to measure eligibility for hire, transfer, promotion, membership, training, referral or retention. This definition includes, but is not restricted to, measures of general intelligence, mental ability and learning ability; specific intellectual abilities; mechanical, clerical and other aptitudes; dexterity and coordination; knowledge and proficiency; occupational and other interests; and attitudes, personality or temperament. The term “test” includes all formal, scored, quantified or standardized techniques of assessing job suitability including, in addition to the above, specific qualifying or disqualifying personal history or background requirements, specific educational or work history requirements, scored interviews, biographical information blanks, interviewers’ rating scales, scored application forms, etc.

**8.1(2)** *“Discrimination” defined.* The use of any test which adversely affects hiring, promotion, transfer or any other employment or membership opportunity of classes protected by Title VII, Civil Rights Act of 1964 and Iowa Code chapter 216 constitutes discrimination unless: The test has been validated and evidences a high degree of utility as described in subrule 8.1(3), and the person giving or acting upon the results of the particular test can demonstrate that alternative suitable hiring, transfer or promotion procedures are unavailable.

**8.1(3)** *Evidence of validity.*

*a.* Each person using tests to select from among candidates for a position or for membership shall have available for inspection evidence that the tests are being used in a manner which does not violate 8.1(2). Such evidence shall be examined for indications of possible discrimination, such as instances of higher rejection rates for minority candidates than nonminority candidates. Furthermore, where technically feasible, a test should be validated for each minority group with which it is used; that is, any differential rejection rates that may exist, based on a test, must be relevant to performance on the jobs in question.

*b.* The term “technically feasible” as used in commission rules means having or obtaining a sufficient number of minority individuals to achieve findings of statistical and practical significance, the opportunity to obtain unbiased job performance criteria, etc. It is the responsibility of the person claiming absence of technical feasibility to positively demonstrate evidence of this absence.

*c.* Evidence of a test’s validity should consist of empirical data demonstrating that the test is predictive of, or significantly correlated with, important elements of work behavior which comprise or are relevant to the job or jobs for which candidates are being evaluated.

(1) If job progression structures and seniority provisions are so established that new employees will probably, within a reasonable period of time and in a great majority of cases, progress to a higher level, it may be considered that candidates are being evaluated for jobs at that higher level. However, where job progression is not so nearly automatic, or the time span is such that higher level jobs or employees’ potential may be expected to change in significant ways, it shall be considered that candidates are being evaluated for a job at or near the entry level. This point is made to underscore the principle that attainment of or performance at a higher level job is a relevant criterion in validating employment tests only when there is a high probability that persons employed will in fact attain that higher level job within a reasonable period of time.

(2) Where a test is to be used in different units of multiunit organization and no significant differences exist between units, jobs, and applicant populations, evidence obtained in one unit may suffice for the others. Similarly, where the validation process requires the collection of data throughout a multiunit organization, evidence of validity specific to each unit may not be required. There may also be instances where evidence of validity is appropriately obtained from more than one company in the same industry. Both in this instance and in the use of data collected throughout a multiunit organization, evidence of validity specific to each unit may not be required: Provided, that no significant differences exist between units, jobs, and applicant populations.

**8.1(4) *Minimum standards for validation.***

a. For the purpose of satisfying the requirements of this chapter, empirical evidence in support of a test's validity must be based on studies employing generally accepted procedures for determining criterion-related validity, such as those described in "Standards for Educational and Psychological Tests and Manuals" published by American Psychological Association, 1200 17th Street, N.W., Washington, D.C. 20036. Evidence of content or construct validity, as defined in that publication, may also be appropriate where criterion-related validity is not feasible. However, evidence for content or construct validity should be accompanied by sufficient information from job analyses to demonstrate the relevance of the content (in the case of job knowledge or proficiency tests) or the construct (in the case of trait measures). Evidence of content validity alone may be acceptable for well-developed tests that consist of suitable samples of the essential knowledge, skills or behavior composing the job in question. The types of knowledge, skills or behavior contemplated here do not include those which can be acquired in a brief orientation to the job.

b. Although any appropriate validation strategy may be used to develop empirical evidence, the following minimum standards, as applicable, must be met in the research approach and in the presentation of results which constitute evidence of validity:

(1) Where a validity study is conducted in which tests are administered to applicants, with criterion data collected later, the sample of subjects must be representative of the normal or typical candidate group for the job or jobs in question. This further assumes that the applicant sample is representative of the minority population available for the job or jobs in question in the local labor market: Where a validity study is conducted in which tests are administered to present employees, the sample must be representative of the minority groups currently included in the applicant population. If it is not technically feasible to include minority employees in validation studies conducted on the present work force, the conduct of a validation study without minority candidates does not relieve any person of that person's subsequent obligation for validation when inclusion of minority candidates becomes technically feasible.

(2) Tests must be administered and scored under controlled and standardized conditions, with proper safeguards to protect the security of test scores and to ensure that scores do not enter into any judgments of employee adequacy that are to be used as criterion measures. Copies of tests and test manuals, including instructions for administration, scoring, and interpretation of test results, that are privately developed and are not available through normal commercial channels must be included as part of the validation evidence.

(3) The work behavior or other criteria of employee adequacy which the test is intended to predict or identify must be fully described; and, additionally, in the case of rating techniques, the appraisal form(s) and instructions to the rater(s) must be included as a part of the validation evidence. Criteria may include measures other than actual work proficiency, such as training time, supervisory ratings, regularity of attendance and tenure. Whatever criteria are used they must represent major or critical work behavior as revealed by careful job analyses.

(4) In view of the possibility of bias inherent in subjective evaluations, supervisory rating techniques should be carefully developed, and the ratings should be closely examined for evidence of bias. In addition, minorities might obtain unfairly low performance criterion scores for reasons other than supervisors' prejudice, as, when as new employees, they have had less opportunity to learn job skills. The general point is that all criteria need to be examined to ensure freedom from factors which would unfairly depress the scores of minority groups.

(5) Differential validity. Data must be generated and results separately reported for minority and nonminority groups wherever technically feasible. Where a minority group is sufficiently large to constitute an identifiable factor in the local labor market, but validation data have not been developed and presented separately for that group, evidence of satisfactory validity based on other groups will be regarded as only provisional compliance with these rules, pending separate validation of the test for the minority group in question. See 8.1(8). A test which is differentially valid may be used in groups for which it is valid but not for those in which it is not valid. In this regard, where a test is valid for two groups but one group characteristically obtains higher test scores than the other without a corresponding

difference in job performance, cutoff scores must be set so as to predict the same probability of job success in both groups.

c. In assessing the utility of a test the following considerations will be applicable:

(1) The relationship between the test and at least one relevant criterion must be statistically significant. This ordinarily means that the relationship should be sufficiently high as to have a probability of no more than 1 to 20 to have occurred by chance. However, the use of a single test as the sole selection device will be scrutinized closely when that test is valid against only one component of job performance.

(2) In addition to statistical significance, the relationship between the test and criterion should have practical significance. The magnitude of the relationship needed for practical significance or usefulness is affected by several factors, including:

1. The larger the proportion of applicants who are hired for or placed on the job, the higher the relationship needs to be in order to be practically useful. Conversely, a relatively low relationship may prove useful when proportionately few job vacancies are available;

2. The larger the proportion of applicants who become satisfactory employees when not selected on the basis of the test, the higher the relationship needs to be between the test and criterion of job success for the test to be practically useful. Conversely, a relatively low relationship may prove useful when proportionately few applicants turn out to be satisfactory;

3. The smaller the economic and human risks involved in hiring an unqualified applicant relative to the risks entailed in rejecting a qualified applicant, the greater the relationship needs to be in order to be practically useful. Conversely, a relatively low relationship may prove useful when the former risks are relatively high.

**8.1(5) *Presentation of validity evidence.*** The presentation of the results of a validation study must include graphic and statistical representations of the relationships between the test and the criteria, permitting judgments of the test's utility in making predictions of future work behavior. (See 8.1(4) "c," concerning assessing utility of a test.) Average scores for all tests and criteria must be reported for all relevant subgroups, including minority and nonminority groups where differential validation is required. Whenever statistical adjustments are made in validity results for less than perfect reliability or for restriction of score range in the test or the criterion, or both, the supporting evidence from the validation study must be presented in detail. Furthermore, for each test that is to be established or continued as an operational employee selection instrument, as a result of the validation study, the minimum acceptable cutoff (passing) score on the test must be reported. It is expected that each operational cutoff score will be reasonable and consistent with normal expectations of proficiency within the work force or group on which the study was conducted.

**8.1(6) *Use of other validity studies.*** In cases where the validity of a test cannot be determined pursuant to 8.1(3) and 8.1(4) (e.g., the number of subjects is less than that required for a technically adequate validation study, or an appropriate criterion measure cannot be developed), evidence from validity studies conducted in other organizations, such as that reported in test manuals and professional literature, may be considered acceptable when:

a. The studies pertain to jobs which are comparable (i.e., have basically the same task elements), and

b. There are no major differences in contextual variables or sample composition which are likely to significantly affect validity.

Any person citing evidence from other validity studies as evidence of test validity for their own jobs must substantiate in detail job comparability and must demonstrate the absence of contextual or sample differences cited in "a" and "b" of this subrule.

**8.1(7) *Assumption of validity.***

a. Under no circumstances will the general reputation of a test, its author or its publisher, or casual reports of test utility be accepted in lieu of evidence of validity. Specifically ruled out are: assumptions of validity based on test names or descriptive labels; all forms of promotional literature; data bearing on the frequency of a test's usage; testimonial statements of sellers, users, or consultants; and other nonempirical or anecdotal accounts of testing practices or testing outcomes.

*b.* Although professional supervision of testing activities may help greatly to ensure technically sound and nondiscriminatory test usage, this alone shall not be regarded as constituting satisfactory evidence of test validity.

**8.1(8)** *Continued use of tests.* Under certain conditions, a person may be permitted to continue the use of a test which is not at the moment fully supported by the required evidence of validity. If, for example, determination of criterion-related validity in a specific setting is practicable and required but not yet obtained, the use of the test may continue, provided: the person can cite substantial evidence of validity as described in 8.1(6); and the person has in progress validation procedures which are designed to produce, within a reasonable time, the additional data required. It is expected also that the person may have to alter or suspend test cutoff scores so that score ranges broad enough to permit the identification of criterion-related validity will be obtained.